

Eberhard Karls University of Tübingen

Course: Practice and Application– Project: Applied English Literatures and Cultures

Instructor: Dr. Christoph Reinfandt

Winter semester 2025-26

Mina Farhadi

Critical Approaches to AI for Humanities Research & Teaching

Workshop Report

Duke University

November, 2025

Contents

Background and Introduction	2
Summary of Presentations and Discussions	3
What is AI	3
Limitations, Ethics, possibilities	4
AI limitations	4
AI Ethics	5
AI Possibilities	6
Responsible Uses of AI in research	6
Brain First approach to the use of AI.....	7
Reflection	7
Bibliography and useful links	9

Background and Introduction

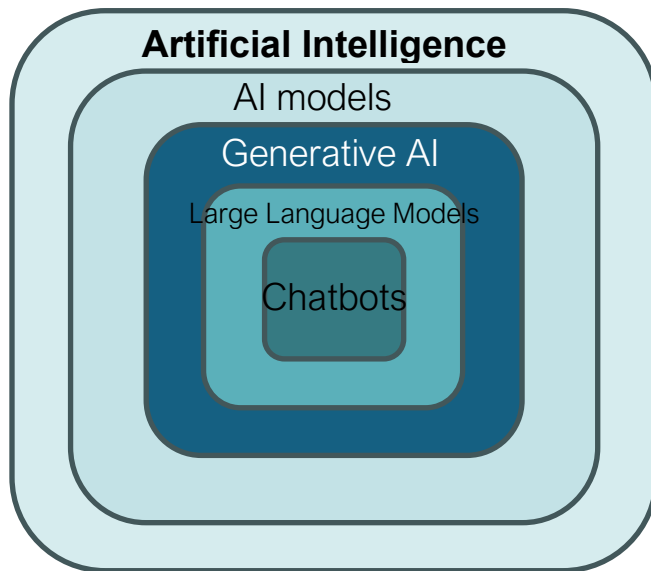
Artificial intelligence has become an undetachable tool in academic research, especially since 2022 when Open AI released Chat GPT. Although the new possibility of submitting cumbersome tasks to AI, opened up new chances to faster and higher quality performance across in many fields, it has also raised unprecedented ethical concerns. In this new era cases such as copyright, privacy protection, preserving the originality of scholarly work, maintaining factual accuracy, and ensuring unbiased information have faced new challenges. While there are some research institutes that are trying to raise awareness about these concerns, they struggle to gain the attention they intend to. Amid the excitement about how AI might revolutionize technology and the competitive race among tech companies to develop faster and more powerful products, the ethical questions, that do not promise immediate economic profit can hardly get the attention they need. To communicate these concerns, many research institutes rely on approaches such as publishing articles, organizing talks, and hosting group discussions. These are intended to raise awareness about the immediate risks of uncritical use of AI and aim to help researchers understand the potential pitfalls. To engage with these initiatives, is an available way to stay informed and to navigate the risks in a rapidly evolving world of AI in research for students and scholars.

This workshop was held online on 4 November 2025, and was facilitated by Duke University Libraries, part of Duke University, United States. The workshop was presented by Hannah Jacobs, a Digital Humanities Specialist at Duke's Digital Art History & Visual Culture Research Lab. It was aimed at students, teachers and researchers from around the world, with the goal of raising awareness about the potential benefits, limitations, and ethical concerns of using AI, in research and teaching. Approximately 50 participants attended via Zoom, most of whom were Duke University students. To engage attendees and gather their opinions, Hannah used Padlet.com, which allowed participants to respond to posed questions while viewing the input of others. The workshop was presented in two sessions. The first session's focus was on engaging participants and highlighting the importance of the topic. It introduced the basics of AI, including its limitations, ethical considerations, and possibilities in application in research. The second session tried to guide participants through practical methods for the responsible use of AI by demonstrating the "Brain First" approach in applying AI to humanities research.

Summary of Presentations and Discussions

What is AI

The workshop began with a simple question of what AI is. After gathering some participants' opinions, the following visualization was presented which shows multiple terms nested within Artificial Intelligence: AI models, Generative AI, Large Language Models (LLMs), and Chatbots. Each term was defined separately.



created by Annie Brown in Duke Libraries

According to Stanford Medicine glossary of artificial intelligence terminology, AI is defined as “a branch of computer science that focuses on creating machines, computer programs, or software systems that can perform tasks typically requiring human intelligence.” Bender and Hanna describe “AI” not as a coherent set of technologies but as technologies focusing on the automation of a specific set of activities, including decision making, classification, recommendation, transcription and translation, and text and image generation. (Bender and Hanna). AI performs these activities via entities that are called models. “Generative AI” was defined as AI models that learn the patterns and structure of their training data and then generate new data with similar characteristics or perform tasks they were not explicitly trained to do. Large Language Models, mostly known as LLMs, are types of generative AI. These systems are trained on string prediction tasks. That is, they predict the likelihood of a token (such as a character, word, or string of characters) based on preceding or surrounding context in the dataset. Finally, interfaces that are designed to engage users and

mimic human conversation through text or voice interactions were introduced as chatbots. Examples of such products include ChatGPT, CoPilot, Gemini, and Claude.

Limitations, Ethics, possibilities

To open up the discussion about Limitations and problems of using AI, findings of an article by Rongqian Ma, Meredith Dedema, and Andrew Cox, was introduced to the attendees. This article examines how Digital Humanities (DH) scholars assess generative AI in their work. Ma et al. use the lens of Actor–Network Theory (ANT) to argue that GenAI is not creating a radical paradigm shift, but it is gradually reshaping the networks that constitute DH practice. The authors of the article shed light on a division in the scholar general opinion about the role of GenAI within DH community. While many acknowledged that AI has improved research processes regarding speed and quantity, they also reported the need for continuous human oversight on AI due to its limitations.

AI limitations

Emily Bender and Alexander Koller in their famous Octopus Paper describe “*meaning* to be the relationship between linguistic form and something external to language” (Bender and Koller p 5187). LLMs cannot achieve understanding only by analyzing forms because they just predict patterns rather than grasping the meaning and their improvement is mostly in getting better at this pattern recognition. When an LLM tries to produce a coherent answer while it does not have access to underlying data or when it is uncertain in pattern prediction, hallucinations arise. When the training patterns used to teach an LLM fail to exactly predict the next token, it may still strive to produce a sound meaning, even at the price of it being nonsensical. In other words, the main issue with LLM hallucinations, is that they prioritize confidence over expressions of uncertainty. Unlike humans who seek for clarifications when a task is not clear enough or when the data at hand to do the task does not suffice, LLMs are programmed to produce a confident response rather than request additional details or admit the lack of sufficient data.

One other barrier in achieving higher level of understanding for LLMs is that they cannot verify facts or assess ambiguity when they lack understanding. These limitations which are inherent to their design significantly increase risks when relying on them without human oversight. IBM defines this phenomenon as a form of AI hallucination in which a generative system “perceives patterns or objects that are nonexistent or imperceptible to human observers, creating outputs that are nonsensical or altogether inaccurate.” (IBM) Although hallucinations and overconfidence were not the only limitations of AI, the workshop could not go to deep analysis of all the technical limitations due to lack of time. Therefore, another major area of concern in the use of AI followed: the ethical questions that arise when these systems are integrated into research and teaching.

AI Ethics

New technologies often challenge established ethical norms, and AI is no exception. Since AI tools, especially chatbots, have become widely accessible and increasingly integrated into research, new ethical concerns emerged. The workshop briefly introduced four of the key ethical issues in this regard: access, labor, privacy and environment.

Access: Technologies always reflect the priorities of those who develop and maintain them. And they mostly aim at bringing economic and social benefits for them. Most of the major tech companies and data centers are located in the Global North, primarily the United States and China. Therefore, there is no wonder that experts from other regions have limited influence on the development, training data, and linguistic coverage of AI systems. The priorities and worries of poorer countries can hardly reach the driving engines of AI and it not only reinforce the existing imbalance of global inequalities but also threatens linguistic diversity, since only a small number of languages are strongly represented in current training datasets.

Labor: Job displacement and the need for learning new skills to be able to remain in an AI driven job market has been widely discussed. Therefore, the workshop tried to highlighted another ethical issue regarding labor: the invisible laborers who support AI systems. Workers, mostly located in economically disadvantaged counties, perform tasks such as reviewing training data for harmful, offensive, or inappropriate material for the AI companies. Although their work is essential to making AI systems function safely, it remains largely hidden from public view and is performed under unfair conditions. There are reports that indicate many of these workers, for example in Kenya or Colombia, are paid very low wages, sometimes even below the average income in their own regions.

Privacy: Maintaining a fully private life is already very difficult with the digitalization in progress. AI's data crawlers can collect and incorporate personal sensitive data in their training dataset usually without obtaining personal knowledge or consent. Companies such as chat GPT that operate AI tools, store user inputs and process them under their own terms and conditions. Therefore, users who include personal data in the prompts unintentionally expose sensitive information about themselves. This can raise serious concerns about consent and data protection.

Environment: Amid all the spectrum of ethical concerns regarding AI implications, environmental consequences have received the least attention. Environmental impacts of AI can be divided to two levels: the environmental cost of individual AI use and the footprints of the infrastructures needed to support AI systems. Estimating energy consumption at personal level per each AI query is difficult due to: the variety of companies, the level of task complexity, the energy source and even the location and climate of data

centers. Sam Altman recently suggested that a single ChatGPT query consumes roughly the same amount of energy as running a high-efficiency light bulb for about two minutes. Another estimate, indicates that each Google Gemini prompt uses as much energy as 9 seconds of television viewing. Considering the billions of prompts submitted to chatbots daily highlights both the inconsistency of these calculations and the vast potential differences in energy usage. The infrastructural environmental issues are also largely overlooked. Rapidly expanding data centers, which are essential for the progress of AI models, require significant amounts of electricity and water for cooling, as well as the burning of methane or other fossil fuels to meet their high and growing energy demands. In many regions these facilities have already caused rise in local energy prices, greenhouse gas emission and has put strain on water supplies. A recent MIT report reveals that, data centers are accountable for an estimate of 4.4% of total use of energy in the US. This paper suggests that by 2028 this number reaches to 22% if it goes in the current trend. (Shehabi et al. 5) This trajectory underscores how AI expansion could outpace sustainable resources available for humans in case left unaddressed.

AI Possibilities

Regardless of all the limitations and ethical concerns that have been mentioned researchers and students do not deter from using AI, because of its massive possibilities. Some of these possibilities include: Promoting inclusive practices in research, restoring and categorizing the old massive archives and digitally restore old films and audios. At this stage of the workshop the attendee got the chance to explain some of the possibilities that AI has brought to their fields, via the Padlet. Brainstorming new ideas, conducting quicker literature review, and assistance in writing were some of the mentioned opportunities that AI has brought to research and teaching areas. Despite all the discussed limitations and ethical concerns, the workshop had a more problem-solving approach and tried to guide users to a more practical and at the same time critical use of AI. Therefore, the next stage of presentation was about simple rules that can be applied while using AI to avoid falling into the aforementioned pitfalls.

Responsible Uses of AI in research

The second session of the workshop began with a presentation of a series of do's and don'ts in responsible use of AI. Prompted methods included: Use of chatbots for summarizing or searching about topics that are already well-known, use of AI tools that actively protect sensitive information, consideration of environmental impact, checking the guideline of the funder or publisher of the project before using AI and finally to use SIFT in asking chatbots in retrieving information. SIFT which is a four-step method that helps in online information evaluation by focusing attention on what matters, was created by Mike Caulfield. These four steps include: **S**top, **I**nvestigate the source, **F**ind better coverage, and **T**race claims to the original context. On the other hand, the workshop strongly recommended scholars to cautious use of AI

in humanities research. In particular participants were advised to avoid citing information found or summarized by a chatbot without authentication, using a chatbot AI before checking its privacy policy, submitting predominantly AI-generated texts as your own work and including AI tools as authors.

Brain First approach to the use of AI

What initially motivated the development of Artificial Intelligence was an effort to model human brains capacity in categorifying and classifying information. Gold in his article “The Brain-Inspired Approach to AI” explains that: “Modern AI research involves creating and implementing algorithms that aim to mimic the neural processes of the human brain. Their goal is to create systems that learn and act in ways similar to human beings.”. (Gold) Therefore, setting aside the role of human brain and assigning tasks that require human interpretation and cognition entirely to AI does not show a rational approach. As the Academic Resource Center of Duke university suggests, that effective use of AI requires humans’ active cognition rather than suspension. A productive perspective on the use of AI in research is to consider it as a tool for extension not a shortcut to human thinking. AI models have neither claimed nor demonstrated the capacity to replace the human’s judgement, planning, and reasoning. Consequently, these are humans who must guide AI system not the other way around. For responsible and meaningful use of AI, the human mind must function as the primary planner, with AI serving as a supporting tool to complete the tasks. This is the perspective that is commonly referred to as “Brain first approach”. To keep up with this approach AI users need to: define and identify the specific use of AI in the tasks, setting preferences by clarifying to AI the best method to undertake an assignment, staying in lead by constant check for accuracy of the results with other reliable source.

At this stage, to reengage the participants the workshop facilitated another discussion using Padlet. The stated question was how the attendees were planning to critically engage with or critically resist the use of AI in teaching and research. Several strategies that were repeatedly mentioned included: Attempting to formulate solutions independently before consulting AI; Verifying the sources of the information that AI has provided us; and posing the same question in multiple ways to assess whether AI generates consistent answers. These were some of the prominent practices that were mentioned as useful approached to foster critical and responsible use of AI.

Reflection

The Critical Approaches to AI for Humanities Research & Teaching workshop was informative and well-structured and could successfully transfer the intended critical view in using AI in the Attendees. Participants had the opportunity to reflect on the input and express their own understanding of them.

Materials were presented with clear coherence, and the stream of information was clear and tangible. One concrete takeaway from the workshop was the list of dos and don'ts in responsible use of AI and brain first approach.

Although the workshop was aimed at researchers and teachers, the content was in fact relevant to all users of AI. Unfortunately, the title could deter other potential attendees, as it seemed something in particular for researchers and teachers, while it could definitely address a wider range of people. One way to boost attendees' engagement quality could be forming break-up groups for discussions. Break up groups usually provide more comfortable and open environment for the participants to engage in the discussions.

The number of talks and seminars and workshops that have critical view to use of AI is increasing and students and other general users of AI must be encouraged in gaining more awareness about the potential risks in using AI irresponsibility, especially when it comes to publishing the result. Moreover, attending events organized by diverse institutions and organizations around the world can be highly valuable, as it exposes participants to a wider range of perspectives and concerns in the rapidly evolving landscape of AI technologies.

Bibliography and useful links

Altman, Sam. “*The Gentle Singularity*.” Blog of Sam Altman, 10 Jun. 2025. blog.samaltman.com/the-gentle-singularity.

Bender, Emily M., and Alexander Koller. *Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data*. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, July 2020, pp. 5185–5198.

Duke University Academic Resource Center. “AI and Learning.” Duke University, arc.duke.edu/ai-and-learning/

Gold, Edem. “The Brain-Inspired Approach to AI – Explained for Developers.” *freeCodeCamp*, <https://www.freecodecamp.org/news/the-brain-inspired-approach-to-ai/>

IBM. “What Are AI Hallucinations?” IBM Think, IBM, 2025 (or most recent update), www.ibm.com/think/topics/ai-hallucinations.

Shehabi, Arman, Sarah J. Smith, Alex Hubbard, Alex Newkirk, Nuoa Lei, Md Abu Bakar Siddik, Billie Holecek, Jonathan Koomey, Eric Masanet, and Dale Sartor. *2024 United States Data Center Energy Usage Report*. Lawrence Berkeley National Laboratory, Dec. 2024. climate.mit.edu/sites/default/files/2025-02/lbnl-2024-united-states-data-center-energy-usage-report.pdf.

Southern New Hampshire University Library. “*Getting Started with Research at Shapiro Library*.” SNHU LibGuides.

Stanford Medicine. “*ABCs of AI: A Brief Glossary of Artificial Intelligence Terminology*.” Stanford Medicine Magazine, 10 Nov. 2023, stanmed.stanford.edu/brief-glossary-artificial-intelligence-ai/.

Zhao, Yachao, Bo Wang, Yan Wang, Dongming Zhao, Ruifang He, and Yuexian Hou. “*Explicit vs. Implicit: Investigating Social Bias in Large Language Models through Self-Reflection*.” arXiv, 4 Jan. 2025.

Useful links

hannah.jacobs@duke.edu

<https://library.duke.edu/>

<https://aclanthology.org/2020.acl-main.463.pdf>

<https://stanmed.stanford.edu/brief-glossary-artificial-intelligence-ai/>

<https://arxiv.org/abs/2501.02295>

<https://www.ibm.com/think/topics/ai-hallucinations>

<https://blog.samaltman.com/the-gentle-singularity>

https://climate.mit.edu/sites/default/files/2025-02/lbnl-2024-united-states-data-center-energy-usage-report.pdf?utm_source=chatgpt.com

<https://libguides.snhu.edu/c.php?g=92303&p=10166599>

<https://arc.duke.edu/ai-and-learning/>